

SYSTEM AND METHOD FOR HIERARCHICAL SEGMENTATION WITH LATENT SEMANTIC INDEXING IN SCALE SPACE

Abstract of the Disclosure

A system and method for automatically generating a hierarchical table of contents or outline for indexing a document and identifying clusters of related information in the document. The document may comprise text, audio, video, or a multimedia presentation. The invention employs a unique and novel combination of latent semantic indexing techniques to identify related blocks and major topic changes within the document with scale space segmentation techniques to respectively identify self-similar blocks within the document and to thus find topic changes of various sizes at block edges. The invention then produces a visual presentation of the semantic structure of the document.